Solving the Documentation Puzzle with Data Analytics

Save to myBoK

By Kapila Monga

"*It's a beautiful day to save lives. Let's have some fun,*" so says Dr. Derek Shepherd (of "Grey's Anatomy" fame) and with a smile on his face, gets on with his day's activities—to help patients heal. I haven't seen many doctors work, except in "Grey's Anatomy," but I have interacted with a couple of them. Being in the analytics field, I have always tried to quiz them on how they do decide which ICD code or which CPT code or which HCPCS code to include. How do they decide which code to pick from among so many codes, the descriptions of which all sound similar to mortals like us?

While all of them have tried to address my questions, they did echo a common sentiment: "Our job is to save lives and help patients heal, not ensure the quality of documentation. We are hard-pressed on time. Apart from treating patients that come to us, we need time to read medical journals, too, and keep ourselves updated with the latest advancements in field. We feel unduly penalized when we are asked to ensure the coding is accurate and exhaustive."

This is a fair point and probably would have been the genesis of the medical transcription industry and the medical coding profession. Given the complexity of our medical system and number and types of codes, and the wealth of data now getting created, this also forces analytics professionals and data scientists to consider how they can help. This blog aims to address this part of the puzzle.

EHR systems have one very interesting piece of information: physician notes. A physician writes these notes for every patient he or she treats. The notes include information on vitals, medications, symptoms he or she has been experiencing, pre-existing conditions, behavioral indicators, the doctor's diagnosis, prognosis and the treatment approach (to name a few), and all this in form of free text. When reading these notes we realize how exhaustive they are—and that they are analogous to the timeline view of a patient's state of body and mind since the last visit.

However, analyzing this information is not easy. It is bytes and bytes of text with no structure, a good degree of subjectivity, and subject to privacy and security regulations. However, the analysis of this information has never been easier than it is today. Use of Big Data analytics technologies to store and process this unstructured data, coupled with the continuous attempt of EHR vendors to bring in more and more structure to this unstructured data, has led to the increased need for analytics professionals and data scientists to mine the information. A very small example of how this information can be useful is presented below for finding missing ICD codes.

We will see below how analysis of physician notes will enable one to answer the question:

Is E11.22, the code for Type 2 Diabetes Mellitus with Diabetic Chronic Kidney Disease, a valid Diagnosis code for the patient?

To make it simple, we will break this in just two steps:

Step 1: Define the bag of words. (According to Wikipedia, the term "bag of words" is a "simplifying representation used in natural language processing.) Physician notes for a patient that exhibits conditions corresponding to diagnosis code E11.22 should ideally have the following words: (Note: Below is a small subset of the actual bag of words, for illustration purposes. The actual list to be used in text mining will have a greater number of terms, and more clinical indicators.)

Chronic kidney + type 2 diabetes mellitus
type 2 DM + Dialysis
DM 2 w diabetic + CKD stage 1
gfr >=90 + Stage 1
gfr 60-89 + w HTN + Stage 2
gfr <15 + w HTN + Stage 5

CKD w HTN
ESRD on dialysis, dialysis w HTN
Hypertension + chronic kidney

Step 2: Use any package or software capable of running text mining on physician notes, and identify physician notes having these bag of words. There will be data cleaning and preparing required, which any data analyst or data scientist can do, followed by mining of text data using relevant procedures and code files. And as a result, you will see the physician notes that have these indicators.

In order to make the algorithm even more accurate, an additional level of validation can be put in place by superimposing information from other fields in other EHR systems over this data. This will help reduce false positives in the detection process described above.

The information present in physician notes can be put to multiple uses, including but not limited to: detecting fraud, waste and abuse; performing disease onset prediction; patient health risk stratification; defining and validating treatment pathways; and medication effectiveness assessments to name a few. Doctors may say that they are not supposed to do documentation, but all they mean is that the documentation may not be written the way users want. After studying the way physicians typically write these notes, I would say we can live with this constraint.

Note: The opinions expressed in this article are the author's own and do not reflect the view of the organization author works for, or of any other corporate entity.

Kapila Monga (kapila.monga@gmail.com) is a Healthcare Analytics professional with 10-plus years of experience across consulting and analytics, for healthcare and life sciences customers. She currently works with Cognizant Technology Solutions in their Healthcare Analytics practice in the US and helps healthcare customers leverage transformative power of analytics and data science to make their business processes more effective.

Original source:
Monga, Kapila. "Solving the Documentation Puzzle with Data Analytics" (Journal of AHIMA website), August 17, 2016.